

Report: 2017 AAHSL Data Management Training Scholarship

I used my scholarship funds to enroll in Springboard's Introduction to Data Science course (<https://www.springboard.com/workshops/data-science>). I selected this course to prepare me in the development of a 4th-year elective for undergraduate medical students, which will expose medical students to advanced data analysis methods. The objectives of the elective are to provide students with an understanding of the underlying statistical principles of various statistical methods and the circumstances in which they are most appropriate. Using R, students will gain hands-on experience in applying these methods to a data set of their choice. The Introduction to Data Science course provided me with valuable practical experience which will inform the development of this elective, particularly the hands-on lab portion of the course.

Springboard's Introduction to Data Science is a 9-unit course including an introduction to R, data wrangling, data analytics, data storytelling, and predictive modeling. Each unit is composed of a set of videos and articles pulled from various online sources, coupled with exercises to provide hands on practice. The core of the course is a data science project that the student selects and works on with the guidance of a mentor. Each week, the student and mentor meet online for 30 minutes to discuss progress in both the course and project. The one-on-one mentoring is extremely helpful, both in understanding the course content and in maintaining steady progress in the course. The course is self-paced and incurs a monthly fee. I completed the course in four months.

My project was to develop predictive models using a de-identified EMR dataset to predict diabetes in patients. The dataset comprised a variety of data for 10,000 patients, some of were identified as having a diabetes diagnosis. Data in the dataset related to diagnosis, medication, allergies, as well as various patient characteristics, measures, and tests. All data directly related to diabetes was removed. I cleaned and worked the data into structures that could be analyzed, performed exploratory analysis on the data, and developed models to identify potential factors that might be used to predict diabetes in patients. I used data visualization to prepare a data story to communicate the results of my analysis in a final report.

In addition to preparing me for the development of this course, I plan to use the knowledge I have gained towards greater involvement in data analysis as a librarian, perhaps eventually leading to a role as a data librarian for the medical library. I still have much to learn, but this has given me a solid foundation for the development of these skills. I am grateful to AAHSL for this scholarship and the opportunity it has provided me in enriching my knowledge base and furthering my career as a medical librarian.

Respectfully,

Keith Engwall, MSLIS, AHIP
Assistant Professor
Web & Emerging Technologies Librarian
Oakland University William Beaumont School of Medicine